

*The correspondence section is a public forum and, as such, is not peer-reviewed. EHP is not responsible for the accuracy, currency, or reliability of personal opinion expressed herein; it is the sole responsibility of the authors. EHP neither endorses nor disputes their published commentary.*

## Hormesis: A Brief Reply to an Advocate

doi:10.1289/ehp.0901681

In his commentary in *Environmental Health Perspectives*, Calabrese (2009) offered a number of responses to my critique of hormesis methodology (Mushak 2009). Here I will provide a counterpoint to that effort.

- Calabrese (2009) falsely asserted that I erred in calculations associated with entry and evaluatory criteria for hormesis frequency, specifically by choosing the wrong denominator for examining the proportion of entry candidates eventually found to be hormetic using the most conventional form of statistical significance. The choice of a denominator for these calculations depends on the question asked. My key question was, What proportions of 668 dose–response entry candidates from 20,285 original articles, using the three criteria identified by Calabrese and Baldwin (2001), partition into each of three hormesis categories? A total of 245 of the 668 candidate dose–responses (37%) had hormetic character, but only 74 of those (30%) were derived using the typical statistical significance test, yielding 11% overall.
- Calabrese (2009) mischaracterized my statements about the reliability of the two unvalidated selection criteria (Mushak 2009). My comments addressed applying criteria to screening large databases of publications for a putative new phenomenon. I was not concerned about routine uses of statistical forms for empirical data (e.g., analyses using 95% confidence intervals on independent means).
- Calabrese (2009) misunderstood my concerns about the two tallies of dosing points (1,089 and 1,791 points) from two of his previous studies (Calabrese and Baldwin 2001, 2003b). The still unanswered question is how the 871 (80% of 1,089) control-equivalent and threshold response-compatible dosing points reported by Calabrese and Baldwin (2001) are mathematically incorporated into a high preponderance of hormetic dosing points (to a 2.5:1 ratio) they reported later (Calabrese and Baldwin 2003b). I was not concerned about simple counts.
- Calabrese misinterpreted my concern about clustered distributions in entry candidates in the 20,285 articles. I was not referring to publications in which the same information is recapitulated in multiple articles, but

whether serial publications that described a given experimental approach but tested different substances were included in the articles database. The clustering pattern, although important, remains unexplained.

- Calabrese stated that the use of entry and evaluation criteria had been validated for both sensitivity and specificity. The question here is whether entry and evaluation criteria that established the original sets of hormetic, false-positive, and false-negative values were validly derived.
- Calabrese (2009) misunderstood and misapplied my rationale for including single sub-NOAEL (no observed adverse effect level) dosing points in the original database. He stated that virtually all of the dosing points within the selected 664 dose responses had been identified previously (Calabrese and Baldwin 2003b). However, in my commentary (Mushak 2009), I clearly conveyed that this step itself had an inherent positive bias and that it is not surprising that hormetic responses outnumbered negative ones. Calabrese was incorrect that including single sub-NOAEL points from the 20,285 articles adds negative bias; rather, such inclusion offsets and corrects an inherent positive bias.
- Calabrese challenged my discussion of the National Cancer Institute (NCI) yeast data set, arguing that the Crump analysis noted in my commentary (Crump 2007) was not peer-reviewed [of course, neither was the rebuttal letter by Calabrese et al. (2007) peer-reviewed]. Calabrese missed the point: Which of two plausible alternatives better addresses the truth of hormesis being present in the NCI data set? Calabrese (2009) noted that Crump's approach introduced 8-fold more variability into the control group statistics, accounting for lack of hormetic evidence. Thereby, he conceded that alleged hormesis in the NCI yeast data lies within the range of determinable control (i.e., nonhormetic) responses.
- Calabrese (2009) challenged my critique of an earlier article on the National Toxicology Program dose-ranging program (Calabrese and Baldwin 2003a). He asserted that all levels of evidence should combine to support the cumulative 31% hormesis frequency. I disagree that poor evidence is just as good as strong evidence; only their "moderate to high" and "high" evidence should have been used in their analysis, yielding a combined 2.3% frequency and not the claimed 31%. The data of Calabrese and Baldwin (2003a)

provided little meaningful support for 31% hormetic frequency.

- Calabrese (2009) objected to my discussing the language issues for hormesis; he argued that (hormesis) revisions are part of the nature of science and new phenomenology, and ignored my point that current hormesis definitions are either those of interpretive convenience or represent divergence rather than convergence (the usual path). One definition in my commentary (Mushak 2009) explained hormesis as an overcompensation for homeostatic preservation; the only discernible basis is as an explanation for U(J)-shaped or inverted U(J)-shaped curves. Another definition explained hormesis as three divergent phenomena.
- Calabrese (2009) took strong exception to my view that public agencies have been slow to address and accommodate hormesis within policy formulations. Regulatory agencies dealing with xenobiotics and human or ecologic health—the key issue—have not adopted hormesis.

*I thank B. Mushak for editing assistance.*

*The author has served as a consultant, advisor, and expert witness over the last 3 years. None of these activities concerned hormesis.*

**Paul Mushak**

PB Associates

Durham, North Carolina

E-mail: [padbmushak@cs.com](mailto:padbmushak@cs.com)

## REFERENCES

- Calabrese EJ. 2009. Hormesis: a conversation with a critic. *Environ Health Perspect* 117:1339–1343.
- Calabrese EJ, Baldwin LA. 2001. The frequency of U-shaped dose responses in the toxicological literature. *Toxicol Sci* 62:330–338.
- Calabrese EJ, Baldwin LA. 2003a. Hormesis at the National Toxicology Program (NTP): evidence of hormetic dose responses in NTP dose-range studies. *Nonlinearity Biol Toxicol Med* 1:455–467.
- Calabrese EJ, Baldwin LA. 2003b. The hormetic dose-response model is more common than the threshold model in toxicology. *Toxicol Sci* 71:246–250.
- Calabrese EJ, Staudenmayer JW, Stanek EJ III, Hoffmann GR. 2007. Hormesis and high throughput studies: Crump's analysis lacks credibility [Letter]. *Toxicol Sci* 98:602–603.
- Crump KS. 2007. Limitations in the National Cancer Institute antitumor drug screening database for evaluating hormesis [Letter]. *Toxicol Sci* 98:599–601.
- Mushak P. 2009. Ad hoc and fast forward: the science of hormesis growth and development. *Environ Health Perspect* 117:1333–1338.

## Hormesis: Calabrese Responds

doi:10.1289/ehp.0901681R

In his letter, Mushak revisits his criticism (Mushak 2009) of previously reported hormesis frequency estimates (Calabrese and Baldwin 2001, 2003; Calabrese et al. 2006, 2008). In my commentary (Calabrese 2009), I addressed and/or rebutted in considerable detail his arguments (Mushak 2009), and no new data require me to revise that

response. Here I address the key areas raised by Mushak's letter, two of which relate to the frequency of hormesis, and the third considers the acceptance of hormesis by the scientific and regulatory communities.

First, a central point of Mushak's commentary (Mushak 2009) and his letter is his assertion that the reported hormesis frequency of 37% (Calabrese and Baldwin 2001) is incorrect and should be 11%. Unfortunately, Mushak used the wrong denominator in his commentary, and he perpetuates this error in his letter. Briefly, we (Calabrese and Baldwin 2001) estimated the frequency of hormesis using *a priori* entry and evaluative criteria; some 668 dose responses satisfied the entry criteria. There were three independent evaluative criteria (i.e., hypothesis testing, nonoverlapping 95% confidence intervals, and alternative quantitative criteria). Of the 668 dose responses, 213 (31.8%) involved hypothesis testing. Of this total, 74 (74/213; 34.7%) satisfied the evaluative criteria for hormesis, a percentage similar to the other two evaluative approaches. When totaled, the three approaches yielded the 37% estimate. Mushak's error is that he used the 74 dose responses that satisfied the evaluative criteria for hypothesis testing not only against the 213 dose responses that had hypothesis testing (which would have been a correct approach) but against all 668 dose responses, even though the remaining 455 dose responses that satisfied the entry criteria lacked hypothesis testing. None of these 455 dose responses could have been evaluated by the statistical criteria. Nonetheless, Mushak combined all the dose responses that satisfied the entry criteria and derived a hormesis frequency based on only dose responses with statistical significance. In so doing, he mistakenly reduced the 37% frequency to 11%. His method is the equivalent of using a raw score for the math component of the Graduate Record Examination (GRE) as the only source of correct answers, and then using all the questions on the math, verbal, and analytic components of the exam as the denominator, even though the student did not take these other components of the test. Such a calculation would give a useless GRE score. His method of hormesis calculation is clearly why he obtained the incorrect lower frequency.

Second, in his letter Mushak continues to cite a letter by Crump (2007) for which there is no support in the literature; also, Crump's letter is based on an assumption about methods that was refuted by the National Cancer Institute investigators who actually did the original work (Calabrese et al. 2007). Mushak apparently does not grasp that Crump's exercise inappropriately introduced 8-fold more variability into the

data analysis. In his letter, Mushak incorrectly and inexplicably claimed that Crump's analysis resulted in my conceding that the hormetic responses that we reported were not different from control responses.

Third, Mushak's inflexibility concerning hormesis is reflected in his comments that minimize the impact of hormesis and its growing applications. Despite the significant biomedical impact of hormesis, Mushak fails to acknowledge the reality that hormetic effects are the basis for how most anxiolytic (Calabrese 2008a), antiepileptic (Calabrese 2008b), memory (Calabrese 2008c; Zoladz and Diamond 2009), Alzheimer disease (Calabrese 2008c; Congdon et al. 2009), and numerous other classes of drugs work (Kastin and Pan 2008; Mattson 2008; Sonneborn 2008; Thong and Maibach 2008), with all such drugs having to pass the regulatory oversight of the Food and Drug Administration for efficacy and safety. On the environmental side, Mushak—in both his letter and his commentary (Mushak 2009)—did not acknowledge that the largest-ever rodent cancer bioassay (24,000 mice) that was designed to determine the nature of the dose response in the low-dose zone for carcinogens revealed hormetic responses for acetyl aminofluorene-induced bladder cancer and that this was affirmed by the 14-member Society of Toxicology expert panel convened to assess these findings (Society of Toxicology ED<sub>01</sub> Task Force 1981). In both his letter and his commentary (Mushak 2009), he also failed to acknowledge that hormesis has had a meteoric rise in recognition and journal citations within the scientific community, with 15 citations per year in the 1980s to > 2,400 in 2009 alone.

On these grounds and those presented in my commentary (Calabrese 2009), I conclude that Mushak's arguments are without merit.

*This effort was sponsored by the Air Force Office of Scientific Research, Air Force Material Command, U.S. Air Force, under grant FA9550-07-0248. The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsement, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. government.*

*The author's host institution, the University of Massachusetts, has received annual financial contributions from ExxonMobil to support low-dose research activities; these contributions were not used to support activities related to this manuscript. The author directs the BELLE project and two annual conferences and obtains funding for these activities from a variety of sources. These funds are processed by the host university. These contributions were also not used to support activities related to this manuscript. During the last*

*3 years he has also received support for travel and honoraria for seminars on hormesis delivered at Lilly and Sanofi-Aventis and several universities.*

**Edward J. Calabrese**

Environmental Health Sciences Division  
School of Public Health and Health Sciences  
University of Massachusetts  
Amherst, Massachusetts  
E-mail: edwardc@schoolph.umass.edu

## REFERENCES

- Calabrese EJ. 2008a. An assessment of anxiolytic drug screening tests: hormetic dose responses predominate. *Crit Rev Toxicol* 38:489–542.
- Calabrese EJ. 2008b. Modulation of the epileptic seizure threshold: implications of biphasic dose responses. *Crit Rev Toxicol* 38:543–556.
- Calabrese EJ. 2008c. Alzheimer's disease drugs: an application of the hormetic dose-response model. *Crit Rev Toxicol* 38:419–452.
- Calabrese EJ. 2009. Hormesis: a conversation with a critic. *Environ Health Perspect* 117:1339–1343.
- Calabrese EJ, Baldwin LA. 2001. The frequency of U-shaped dose response in the toxicological literature. *Toxicol Sci* 62:330–338.
- Calabrese EJ, Baldwin LA. 2003. The hormetic dose-response model is more common than the threshold model in toxicology. *Toxicol Sci* 71:246–250.
- Calabrese EJ, Stanek EJ, Nascarella M, Hoffmann G. 2008. Hormesis predicts low-dose responses better than threshold models. *Inter J Toxicol* 27:369–378.
- Calabrese EJ, Staudenmayer JW, Stanek EJ III, Hoffmann G. 2006. Hormesis outperforms threshold model in National Cancer Institute antitumor drug screening database. *Toxicol Sci* 94:368–378.
- Calabrese EJ, Staudenmayer JW, Stanek EJ III, Hoffmann GR. 2007. Hormesis and high throughput studies: Crump's analysis lacks credibility [Letter]. *Toxicol Sci* 98:602–603.
- Congdon EE, Figueroa YH, Wang L, Toneya G, Chang E, Kurek J, et al. 2009. Inhibition of tau polymerization with a cyanine dye in two distinct model systems. *J Biol Chem* 284:20830–20839.
- Crump KS. 2007. Limitations in the National Cancer Institute antitumor drug screening database for evaluating hormesis [Letter]. *Toxicol Sci* 98:599–601.
- Kastin AJ, Pan W. 2008. Peptides and hormesis. *Crit Rev Toxicol* 38:629–631.
- Mattson MP. 2008. Awareness of hormesis will enhance future research in basic and applied neuroscience. *Crit Rev Toxicol* 38:633–639.
- Mushak P. 2009. Ad hoc and fast forward: the science of hormesis growth and development. *Environ Health Perspect* 117:1333–1338.
- Society of Toxicology ED<sub>01</sub> Task Force. 1981. Re-examination of the ED<sub>01</sub> study. Overview. *Fundam Appl Toxicol* 1:28–63.
- Sonneborn JS. 2008. Hormetic triggers for intervention in aging, disease and trauma. *Amer J Pharmacol Toxicol* 3:4–13.
- Thong HY, Maibach HI. 2008. Hormesis [biological effects of low level exposure (BELLE)] and dermatology. *Dose Response* 6:1–15.
- Zoladz PR, Diamond DM. 2009. Linear and non-linear dose-response functions reveal a hormetic relationship between stress and learning. *Dose Response* 7:132–148.

## Lead in Drinking Water as a Public Health Challenge

doi:10.1289/ehp.1001979

In drinking water supplies the intake of the toxic heavy metal lead is commonly due to metal corrosion in the peripheral water distribution system, especially the user's plumbing or lead service lines. Recently, the problem again received attention in the United States when testing data of drinking water

at schools was published (Renner 2009). In Europe several countries are known to have significant numbers of buildings with elevated lead tap water concentrations, for example, the United Kingdom (Watt et al. 1996), Austria (Haider et al. 2002) and Germany (Becker et al. 2001).

Lead exposure from drinking water has been a topic of public health prevention programs in several parts of Germany before, for example, Hamburg (Fertmann et al. 2004) and Frankfurt (Hentschel et al. 1999). In 2005 in the northern German state of Lower Saxony, a prevention program was initiated comprising three different approaches at the same time to achieve a widespread effect. To assess the present state of drinking water contamination with lead, a free examination of lead in tap water (after nocturnal stagnation) was offered in cooperation with local public health departments for private households that included young women and families with children (Zietz et al. 2007, 2009). Along with the collection of data, the program aimed to focus public attention on this public health problem. In another part of this program, data from local public health departments on existing lead measurements, especially in public buildings, were collected and analyzed (Zietz et al. 2007, 2009). Finally, a working group on lead replacement, consisting of representatives of all relevant parties (e.g., tenant and landlord associations, crafts people, building and health administrations) was initiated. In the screening part of the project, a total of 2,901 tap water samples from households were collected during 2005–2007. Of these, 7.5% had lead concentrations > 10 µg/L (recommended limit of the World Health Organization) and 3.3% had concentrations above the present limit of the German drinking water ordinance (25 µg/L) (Zietz et al. 2009). We found remarkable regional differences in the frequency of tap water contamination. An additional inclusion criterion in this study was that buildings must have been constructed before 1974 (after which no new lead pipes were installed); therefore, the results cannot be compared directly to other studies. From the data, we roughly estimated that about 4.7% of all households in Lower Saxony have lead concentrations > 10 µg/L (Zietz et al. 2009). In an earlier study in southern Lower Saxony (Zietz et al. 2001a), households with mothers of newborn babies from the area around the university city of Göttingen were investigated. Of the 1,434 stagnation samples, 3.1% had lead concentrations > 10 µg/L.

A moderately higher percentage of households with elevated composite water samples was found in the geographic area of the city of Berlin using two composite water sampling

methods (5.6% and 7.0%, respectively). In total, 2,109 households were tested with both methods in the federal state of Berlin (Zietz et al. 2001b). In a representative study of samples collected in all parts of Germany during 1997–1999 (Becker et al. 2001), the 90th percentile of lead concentrations in 4,761 stagnation samples was 7.6 µg/L.

Projects in association with epidemiologic investigations also provide an opportunity to design prevention programs in this field. Generally, we favor the precautionary measure of preventing exposure to lead by replacing pipes completely. The addition of anticorrosive substances to the public water supply can be effective in lowering lead concentrations. In contrast, changing water chemistry (e.g., a new water disinfectant method, as in Washington, DC, USA) can have a substantial effect in elevating lead (Renner 2009). Flushing the water pipes and using only cold water are short-term methods of decreasing exposure to lead from tap water. Using bench-top water filters can also decrease lead concentrations, but problems such as leaching of different substances into the water or microbial contamination may arise under certain conditions. Thus, lead plumbing material in buildings still poses a challenge for public health in the United States and in Europe.

*The authors declare they have no competing financial interests.*

**Björn P. Zietz**

**Jessica Laß**

**Roland Suchenwirth**

Governmental Institute of Public Health  
of Lower Saxony (Niedersächsisches  
Landesgesundheitsamt)  
Hannover, Germany  
E-mail: bjoern.zietz@nlga.niedersachsen.de

**Hartmut Dunkelberg**

Medical Institute of General Hygiene and  
Environmental Health  
University of Göttingen  
Göttingen, Germany

## REFERENCES

- Becker K, Kaus S, Helm D, Krause C, Meyer E, Schulz C, Seiwert M. 2001. Umwelt-Survey – 1998, Band IV: Trinkwasser – Elementgehalte in Stagnationsproben des häuslichen Trinkwassers der Bevölkerung in Deutschland [in German]. Berlin/WaBoLu-Hefte, Umweltbundesamt.
- Fertmann R, Hentschel S, Dengler D, Janssen U, Lommel A. 2004. Lead exposure by drinking water: an epidemiological study in Hamburg, Germany. *Int J Hyg Environ Health* 207:235–244.
- Haider T, Haider M, Wruss W, Sommer R, Kundi M. 2002. Lead in drinking water of Vienna in comparison to other European countries and accordance with recent guidelines. *Int J Hyg Environ Health* 205:399–403.
- Hentschel W, Karius A, Heudorf U. 1999. Stadtgesundheitsamt Frankfurt. Das Frankfurter Bleiprojekt Maßnahmen zur Einhaltung des Grenzwertes für Blei im Trinkwasser [in German]. *Bundesgesundheitsbl* 42:902–910.
- Renner R. 2009. Out of plumb: when water treatment causes lead contamination. *Environ Health Perspect* 117:A542–A547.
- Watt GC, Britton A, Gilmour WH, Moore MR, Murray GD, Robertson SJ, et al. 1996. Is lead in tap water still a public health problem? An observational study in Glasgow. *BMJ* 313:979–981.
- Zietz B, Dassel de Vergara J, Kevekordes S, Dunkelberg H. 2001a. Lead contamination in tap water of households with children in Lower Saxony, Germany. *Sci Total Environ* 275:19–26.
- Zietz BP, Laß J, Dunkelberg H, Suchenwirth R. 2009. Die Bleibelastung des niedersächsischen Trinkwassers bedingt durch Korrosion von Rohrleitungsmaterialien [in German]. *Gesundheitswesen* 71:265–274.
- Zietz BP, Laß J, Suchenwirth R. 2007. Assessment and management of tap water lead contamination in Lower Saxony, Germany. *Int J Environ Health Res* 17:407–418.
- Zietz BP, Paufler P, Keßler-Gaedtke B, Dunkelberg H. 2001b. Bleiverunreinigung von Trinkwasser bedingt durch Leitungssysteme in Berlin [in German]. *UWSF - Z Umweltchem Ökotox* 13:153–157.

*Editor's note: A second feature by Rebecca Renner in the February 2010 issue of EHP [Environ Health Perspect 118:A68–A72] further explored tap water as a source of potential lead exposure. A third feature in the May 2010 issue will address the public health implications of partial replacement of lead service lines.*